

# Efficient design of cluster randomized and multicentre trials with unknown intraclass correlation

Citation for published version (APA):

van Breukelen, G. J. P., & Candel, M. J. J. M. (2015). Efficient design of cluster randomized and multicentre trials with unknown intraclass correlation. *Statistical Methods in Medical Research*, 24(5), 540-556. <https://doi.org/10.1177/0962280211421344>

**Document status and date:**

Published: 01/10/2015

**DOI:**

[10.1177/0962280211421344](https://doi.org/10.1177/0962280211421344)

**Document Version:**

Publisher's PDF, also known as Version of record

**Document license:**

Taverne

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Download date: 05 May. 2023

# Efficient design of cluster randomized and multicentre trials with unknown intraclass correlation

Gerard JP van Breukelen and Math JJM Candel

Statistical Methods in Medical Research  
2015, Vol. 24(5) 540–556

© The Author(s) 2011

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0962280211421344

smm.sagepub.com



## Abstract

For cluster randomized and multicentre trials evaluating the effect of a treatment on persons nested within clusters, equations have been published to compute the optimal sample sizes at the cluster and person level as a function of sampling costs and intraclass correlation (*ICC*). Here, optimal means maximum power and precision for a given sampling budget, or minimum sampling costs for a given power and precision. However, the *ICC* is usually unknown, and the optimal sample sizes depend strongly on this *ICC*. To overcome this local optimality problem, this study presents Maximin designs (MMDs) based on relative efficiency (RE) and efficiency. These designs perform well over a range of possible *ICC* values either in terms of RE compared with the locally optimal designs, or in terms of minimum efficiency (maximum variance) of the treatment effect estimator. The use of MMDs is illustrated using information from many cluster randomized trials in primary care. It is concluded that MMDs and the optimal design for an *ICC* halfway its assumed range are efficient for a range of *ICC* values and recommendable for practical use. This requires that trial reports mention the study cost per cluster and person.

## Keywords

Cluster randomized trials, cost effectiveness, efficiency, intraclass correlation, Maximin design, multicentre trials, optimal design, power, sample size

## 1 Introduction

To evaluate the effect of a treatment on persons nested within clusters, for instance patients in general practices or pupils in schools, one can randomly assign treatments to individuals within clusters (multicentre trial) or to entire clusters (cluster randomized trial). Cluster randomized trials are performed when randomization of individuals is logistically impossible or may lead to serious treatment contamination. Examples are the evaluation of a smoking prevention program for school children where the program consists of lessons in the classroom<sup>1</sup> and patient-centred care of

Department of Methodology and Statistics, Maastricht University, Maastricht, The Netherlands.

### Corresponding author:

Gerard JP van Breukelen, Department of Methodology and Statistics, School for Public Health and Primary Care CAPHRI, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

Email: gerard.vbreukelen@maastrichtuniversity.nl

newly diagnosed diabetes in general practices.<sup>2</sup> Data analysis has to take the nesting of persons within clusters into account, since ignoring this leads to underestimation of the standard error of the treatment effect estimate, and thus to a too large type I error risk and a too narrow confidence interval. The best method is mixed regression, also known as multilevel analysis, but an unpaired *t*-test on cluster means can be acceptable. The latter is even equivalent to mixed regression if the number of persons sampled is the same for each cluster, the outcome is quantitative and there are no covariates.<sup>3</sup>

Multicentre trials differ from cluster randomized trials in that individuals are randomized within each centre, which is more efficient but not always possible and more prone to treatment contamination. Again, the nesting has to be taken into account, either by mixed regression or by a paired *t*-test of treated *versus* control with centre as unit of analysis, assuming centre is considered a random effect. The alternative of centre as fixed effect is sometimes a better choice, depending on various considerations including the number of centres in the study,<sup>4</sup> but will not be covered in this article.

The planning of both types of trials involves several design choices, including the sample size per level, that is, the number of clusters and the number of persons per cluster. This article is about sample size planning in cluster randomized and multicentre trials. The optimal sample size, henceforth called the optimal design, can be defined in either of two ways, depending on the preconditions. If the total budget for sampling and measuring clusters and persons is fixed, then the optimal design is that which maximizes the precision of the treatment effect estimate and the power of its significance test. If the required power for testing and precision for estimation of the treatment effect are fixed, then the optimal design is that which minimizes the total costs for sampling and measuring clusters and persons. Optimal trial design is of importance from the viewpoint of cost effectiveness, a criterion by which not only medical treatments, but also scientific studies can be evaluated. After all, improving cost effectiveness of trials allows more trials to be run with the same budget. Formulae for computing the optimal design under either of both preconditions, power or budget, are available,<sup>5–11</sup> but make two restrictive assumptions: (1) the number of persons sampled is the same for each cluster and (2) the intraclass correlation (*ICC*) is known, where *ICC* is the proportion of outcome variance that is between clusters rather than within clusters.

These assumptions are not very plausible, as acknowledged in a recent paper on statistical issues in trial design.<sup>12</sup> Although equal cluster sizes are optimal,<sup>13,14</sup> they are rarely feasible. Variation of the actual size of general practices or schools, but also non-response and dropout of persons, lead to unequal cluster sizes in the study and thereby to a loss of power. Likewise, the *ICC* is unknown before data analysis, although systematic reviews suggest that it rarely exceeds 0.10.<sup>2,12,15–17</sup> This is a problem since the optimal design depends on the unknown *ICC* value. In optimal design literature, this is known as the local optimality problem.<sup>18,19</sup> The problem of cluster size variation was addressed by several authors who proposed a sample size correction based on the coefficient of variation (*CV*) of cluster sizes.<sup>14,15,20–26</sup> This article addresses the problem of an unknown *ICC*, by deriving Maximin designs (MMDs) which are either highly efficient compared with the locally optimal designs (LODs) for a range of *ICC* values (maximin relative efficiency), or highly efficient in the worst case scenario of a high *ICC* (maximin efficiency). MMD has been applied to longitudinal studies<sup>27,28</sup> and functional magnetic resonance imaging,<sup>29</sup> among others, but not yet to cluster randomized and multicentre trials. The outline of this article is as follows. Section 2 summarizes the design and *ICC* values from a large number of cluster randomized trials in primary care, showing the wide variation in cluster size and *ICC* as motivation for this study. Section 3 presents the mixed effects model and current sample size formulae for cluster randomized trials with a quantitative outcome, showing how the optimal design depends on the *ICC* and is thus locally optimal only. Section 4 shows how an MMD

overcomes this dependence, and Section 5 compares the efficiency of MMDs with some simple alternatives. Sections 6 and 7 show how the theory also applies to multicentre trials with treatment by centre interaction and to binary outcomes. Section 8 compares the actual designs of Section 2 with LODs and MMDs. The final section summarizes the main results and points out some limitations addressed in ongoing work.

## 2 Designs and ICCs in published cluster randomized trials

A systematic review by Adams *et al.*<sup>16</sup> provides information on the sample sizes and ICCs in 19 cluster randomized trials and 12 other nested designs in primary care in the UK published between 2000 and 2002. Summaries of 152 trials worldwide and published between 1997 and 2000 are given by Eldridge *et al.*<sup>17</sup> Additional data were obtained from six trials in five papers focussing on ICC values from specific trials which were not covered by the two reviews.<sup>30–34</sup> Table 1 summarizes the sample sizes and ICCs found in both reviews,<sup>16,17</sup> showing a large variation between trials in sample sizes and ICCs. The mean cluster size varied from 2 to 151 or from 9 to 82. The ICC ranged from 0 to 0.10 or from 0 to 0.21. The data from six further trials gave similar results and are not further discussed here. We now turn to optimal and robust sample sizes as a function of the ICC and sampling costs, coming back to Table 1 in Section 8.

## 3 Cluster randomized trials: model and optimal sample size

Suppose that a new treatment is evaluated by sampling  $K$  clusters (general practices, schools, etc.) of  $n$  persons (patients, pupils), followed by randomized treatment assignment to clusters. Let the effectiveness of the treatment be expressed as the mean difference between treated and control on a quantitative outcome  $Y$ . The model for data analysis is then:

$$Y_{ij} = \beta_0 + \beta_1 X_j + u_{0j} + e_{ij} \quad (1)$$

where  $Y_{ij}$  is the outcome for person  $i$  in cluster  $j$ , and  $X_j$  the treatment assigned to cluster  $j$  and will be coded 1 for treated and 0 for control clusters. Finally,  $u_{0j}$  is a random cluster effect and  $e_{ij}$  a residual error reflecting person and measurement error effects. These two random effects are assumed to be independently and normally distributed with variances  $\sigma_0^2$  and  $\sigma_e^2$ . The outcomes  $Y_{ij}$  and  $Y_{i'j}$  of any two persons  $i$  and  $i'$  in an arbitrary cluster  $j$  are correlated, and this ICC equals

$$\rho = \frac{\sigma_0^2}{\sigma_Y^2} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_e^2} \quad (2)$$

where the denominator is the total outcome variance within each treatment condition.

The variance of the Maximum Likelihood estimator of treatment effect  $\beta_1$  in model (1) is:<sup>7,35</sup>

$$\text{Var}(\hat{\beta}_1) = [(n-1)\rho + 1] \times \frac{4\sigma_Y^2}{nK} \quad (3)$$

where  $n$  and  $K$  are the number of persons per cluster and the number of clusters so that  $nK$  is the total sample size,  $\rho$  the ICC defined in (2), and  $\sigma_Y^2$  the total outcome variance. The factor  $[(n-1)\rho + 1]$  in (3) is known as the design effect. If  $\rho = 0$ , the design effect is 1 and  $\text{Var}(\hat{\beta}_1)$  for cluster randomization is equal to  $\text{Var}(\hat{\beta}_1)$  for individual randomization.

**Table 1.** Summary of sample sizes and ICCs found in systematic reviews of cluster randomized trials in primary care in the UK between 1997 and 2002 (IQR = interquartile range).

Review	Number of studies	Number of clusters per study	Mean cluster size per study	ICC
Adams <i>et al.</i> <sup>16a</sup>	All 31 studies	Median = 31 IQR = 20–57 Range = 8–158	Median = 23 IQR = 9–75 Range = 2–151	Median = 0.011 IQR = 0–0.024 Range = 0–0.099 90 percentile = 0.046 95 percentile = 0.075
Adams <i>et al.</i> <sup>16</sup>	Only the 19 cluster randomized trials	Median = 41 IQR = 24–64 Range = 12–158	Median = 23 IQR = 6–81 Range = 2–151	Median = 0.011 IQR = 0–0.024 Range = 0–0.099 90 percentile = 0.046 95 percentile = 0.099
Eldridge <i>et al.</i> <sup>17b</sup>	71 for sample sizes, and 13 for ICCs	Median = 34 Range = 2–810	Median = 32 IQR = 9–82	Median = 0.04 IQR = 0–0.21

Notes: <sup>a</sup>The review covered 31 studies with a nested design, of which 19 were cluster randomized, 3 were multicentre, and 9 were surveys. ICCs based on Table 1 in Adams *et al.*<sup>16</sup> and taking per study the median ICC in case of multiple outcomes.<sup>16</sup>

<sup>b</sup>Sample size information was available from 71 of 152 trials, ICC information from 13 trials, first ICC quartile reported to be –0.02, here truncated to 0.

To find the optimal design a cost function must be defined. Assume that inclusion of a cluster into the study costs  $c$  currency units (e.g. Euros, GBP or USD), and that inclusion of a person (subject) within an already included cluster costs  $s$  units, with  $c > s$  usually. The total budget  $B$  needed for including  $K$  clusters of  $n$  persons each, excluding all study costs that do not depend on the sample size at either level, then is:

$$B = cK + snK = K(c + sn) \quad (4)$$

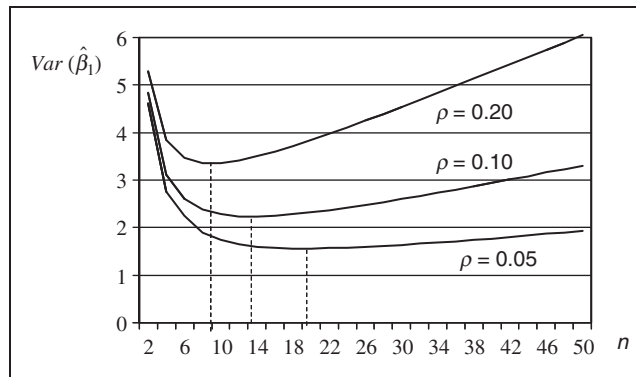
where the factor  $(c + sn)$  is the total sampling cost per cluster of size  $n$ . Finding the optimal design means finding the  $K$  and  $n$  which minimize  $\text{Var}(\hat{\beta}_1)$  in (3) given the constraint in (4). It has been shown<sup>5,7</sup> that this gives as optimal design:

$$K = \frac{B}{\sqrt{\theta sc} + c}, \quad n = \sqrt{\frac{\theta c}{s}}$$

where

$$\theta = \frac{1 - \rho}{\rho} \quad (5)$$

This shows that as the ICC  $\rho$  increases (so that  $\theta$  decreases), the number of clusters  $K$  must increase, while the number of persons per cluster  $n$  must decrease.



**Figure 1.** Variance of the treatment effect estimator as a function of cluster size  $n$  (and implicitly also the number of clusters  $K = B/(c + sn)$ ), for several values of the ICC  $\rho$ , assuming a sampling budget  $B = 100,000$ , sampling costs  $c = 2000$  and  $s = 100$ , and total outcome variance  $\sigma_Y^2 = 100$ . Dotted lines indicate the LOD for each ICC value.

Inserting (5) in (3) gives the variance of the treatment effect estimator for the optimal design, which is the smallest possible variance given the budget  $B$ , the costs  $c$  and  $s$ , the total outcome variance, and the ICC:

$$\text{Var}(\hat{\beta}_1) = g(\rho) \times \frac{4\sigma_Y^2}{B}, \text{ where } g(\rho) = \left( \sqrt{\rho c} + \sqrt{(1-\rho)s} \right)^2 \quad (6)$$

Figure 1 shows how this minimal  $\text{Var}(\hat{\beta}_1)$  depends on the ICC and the cluster size  $n$ , and implicitly on the number of clusters  $K$  due to the constraint in (4). Hence, the power and precision for treatment effect evaluation increase with budget  $B$  for sampling, and decrease with sampling costs  $c$  and  $s$ , in a way which depends on the ICC. More specifically,  $g(\rho)$  runs from  $s$  if  $\rho = 0$  to  $c$  if  $\rho = 1$ , with a maximum  $g(\rho) = c + s$ , attained if  $\rho = \frac{c}{c+s}$ .

There are some limitations to the application of (5). Apart from the fact that the ICC is unknown, which is addressed in the next section, a problem is that there are often boundaries on  $K$  or  $n$ . For example, in a trial on general practice care, only 20 practices may be within reachable distance and willing to participate. Likewise, in a smoking prevention program for school children which is delivered in the class room, whole classes will be included and so the number of included pupils per school may be a multiple of class size. One can then choose from all feasible designs the one which is closest to the optimal design in terms of  $K$  and  $n$  and adapt either  $K$  or  $n$  (whichever can be adapted) to approximate the power or the budget for the optimal design.

#### 4 Unknown intraclass correlation: MMD

Computing the optimal design with (5) requires knowledge or a guess of the ICC, and misspecification may lead to an inefficient design with too many or too few clusters. In optimal design literature,<sup>18,19</sup> this is called the local optimality problem, since a design which is optimal for some ICC value is not optimal for other values within the possible range  $[a, b]$  (called the parameter space), and so the design is locally optimal only. One solution is MMD.<sup>19,27–29</sup> Finding MMD involves three steps:

- (1) First, define the parameter and the design spaces. For example, the *ICC* may be between  $a = 0.01$  and  $b = 0.10$  (parameter space) based on previous studies, and the number of clusters  $K$  may be between 4 and 80, and cluster size  $n$  may be between 10 and 100 (design space), based on feasibility constraints.
- (2) For each *ICC* value in the parameter space, first compute the LOD with (5). Then, compute the relative efficiency (*RE*) of each design in the design space compared with the LOD, by taking the ratio of Equation (6) to (3).
- (3) For each design in the design space, compute its minimum *RE*, that is, find its smallest *RE* value within the parameter space. Then, select that design which maximizes the minimum *RE*, that is, which has the highest minimum *RE* among all designs in the design space. This is the MMD and its minimum *RE* is called the Maximin value (MMV).

These steps give the design which is most robust against misspecification of the *ICC* and thus most cost-effective in the following sense: The inverse of the *RE* of a design *versus* the LOD gives the ratio of the  $\text{Var}(\hat{\beta}_1)$ 's under both designs for the same budget  $B$ , and thereby also the ratio of the budgets  $B$  needed for the designs to have the same power and precision. Hence, a design with a minimum *RE* of 0.80 never needs more than  $1/0.8 = 1.25$  times as much budget as the LOD, no matter what the *ICC* is. By maximizing the minimum *RE*, the MMD follows the optimal design as closely as possible over the whole *ICC* range, and minimizes the worst-case budget increase relative to the optimal design.

Instead of using the *RE*, MMD can also use the efficiency as criterion. Maximizing not the minimum *RE*, but the minimum efficiency itself, is equivalent to choosing the design which minimizes the maximum  $\text{Var}(\hat{\beta}_1)$ . This gives as solution the LOD for the maximum possible *ICC* (which is  $b$ ), since  $\text{Var}(\hat{\beta}_1)$  is an increasing function of the *ICC* for any given design  $(n, K)$  (Equation (3)). This design, denoted as  $\text{LOD}(b)$ , is safe in considering the worst-case scenario of a high *ICC*. However, it is highly dependent on the chosen *ICC* range and very inefficient if the *ICC* is close to its lower boundary, which appears to occur quite frequently according to the *ICC* reviews.<sup>16,17</sup> Moreover, the number of clusters increases with the *ICC* (Equation (5)) and so the  $\text{LOD}(b)$  may not be feasible. In contrast, the MMD leads to a design that is optimal for some *ICC* value well within the *ICC* range, while maximizing the *RE* for other *ICC* values, in particular for the *ICC* boundaries as we will see. Also, the MMD requires fewer, though larger, clusters than the  $\text{LOD}(b)$ . Both types of MMD, based on the *RE* and based on efficiency, are found in the literature and will be compared in this article. Two alternatives will also be compared with the MMDs: the  $\text{LOD}(a)$  design, which is the LOD design for the smallest possible *ICC*, and the  $\text{LOD}(ab)$  design, which is the LOD design halfway the *ICC* range, that is, at  $\text{ICC} = (a + b)/2$ . The  $\text{LOD}(a)$  minimizes the minimum  $\text{Var}(\hat{\beta}_1)$  and is thus an optimistic design. The  $\text{LOD}(ab)$  minimizes the expected  $\text{Var}(\hat{\beta}_1)$  if the *ICC* is symmetrically distributed over its range  $[a, b]$ .

As the  $\text{LOD}(ab)$  already suggests, other solutions to the local optimality problem than Maximin are possible. Using a Bayesian approach, one may maximize the expected instead of the minimum efficiency, by integrating over a probability distribution for the *ICC*,<sup>36–39</sup> thereby taking into account that some *ICC* values within the range are more plausible than others. Or one can minimize the expected instead of maximum  $\text{Var}(\hat{\beta}_1)$ , giving the  $\text{LOD}(ab)$  in case of a symmetric *ICC* distribution. These two Bayesian criteria will not, in general, give the same design, since efficiency and  $\text{Var}(\hat{\beta}_1)$  are non-linearly related. Also, the choice of the *ICC* probability distribution affects the optimal design. Therefore, Bayesian design is beyond the present scope, apart from the special case of the  $\text{LOD}(ab)$ . So, let us look at the MMD based on relative efficiency (MMD) now, ignoring at first any constraints on the design space apart from  $n, K > 0$ . Constraints are discussed at the end of this section.



Using Equations (3) and (6), we get the following expression for the relative efficiency of a given design  $(n, K)$  compared with the LOD( $\rho$ ), as a function of  $n$  and  $ICC$ :

$$RE(n, \rho) = \frac{g(\rho)}{[(n-1)\rho + 1]} \times \frac{n}{c + sn} \quad (7)$$

where  $g(\rho) = (\sqrt{\rho c} + \sqrt{(1-\rho)s})^2$  as in (6), and  $K$  and  $B$  cancel out due to (4), making the  $RE$  a function of cluster size  $n$  and  $ICC$   $\rho$  only, given the costs  $c$  per cluster and  $s$  per person. The following can now be shown (Appendix): If the parameter space for the  $ICC$  is the interval  $[a, b]$ , where  $0 \leq a < b \leq 1$ , then MMD, which maximizes the minimum of the  $RE$  in Equation (7), is

$$MMD: \quad n = \frac{(b-1)g(a) - (a-1)g(b)}{bg(a) - ag(b)}, \quad K = \frac{B}{c + sn} \quad (8)$$

where  $g(\rho)$  is as defined in (6), and  $K$  follows from  $n$  in (8), using (4). By dividing  $g(a)$  and  $g(b)$  by the sampling cost per person  $s$ , it follows that the cluster size  $n$  of the MMD depends on the  $ICC$  boundaries  $a$  and  $b$  and cost ratio  $c/s$  only. The costs  $c$  and  $s$  themselves and the budget  $B$  do not affect the cluster size  $n$ , but the number of clusters  $K$ , of the MMD.

Further, the MMD attains its minimum  $RE$  or MMV at  $\rho = a$  and at  $\rho = b$ , the two minima being equal for the MMD (Appendix), and this MMV is (Equation (7)):

$$MMV = \frac{ng(a)}{(na + (1-a))(c + sn)}, \quad n = MMD \quad (9)$$

which just like the cluster size  $n$  for the MMD, depends on  $a$  and  $b$  and cost ratio  $c/s$  only, as can be verified by dividing numerator and denominator of (9) by the cost per person  $s$ .

As a special case, if  $a = 0$  then  $g(a) = s$ , and we obtain as MMD

$$MMD: n = \frac{g(b)}{bs} + \left(\frac{b-1}{b}\right) = \frac{c}{s} + 2\sqrt{\left(\frac{c}{s}\right)\left(\frac{1-b}{b}\right)} \quad (10)$$

where the term  $\sqrt{(\dots)}$  is the LOD( $b$ ) (Equation (5)). Hence, if  $a = 0$ , then MMD has a cluster size  $n$  equal to the cost ratio  $c/s$  plus twice the cluster size of the LOD ( $b$ ) design.

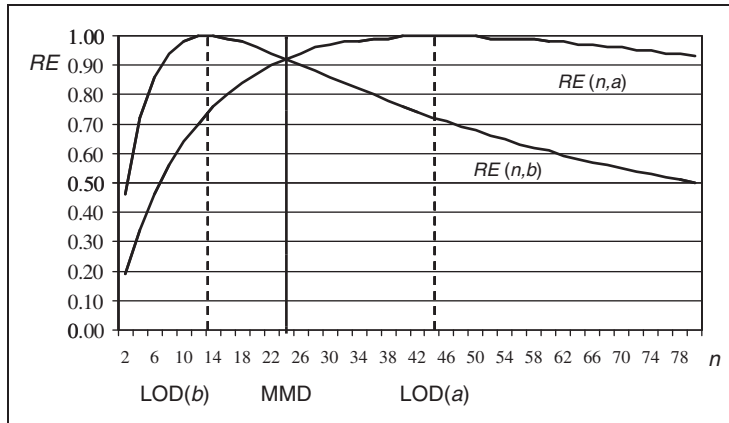
Consequently, the MMD cluster size increases with the cost ratio, but decreases as the maximum possible  $ICC$  value  $b$  increases, reaching a minimum of  $n = c/s$  if  $b = 1$ . In that case, half of the budget is spent on clusters and half on persons (Equation (4)).

Further, if  $a = 0$ , we get as minimum  $RE$  of the MMD:

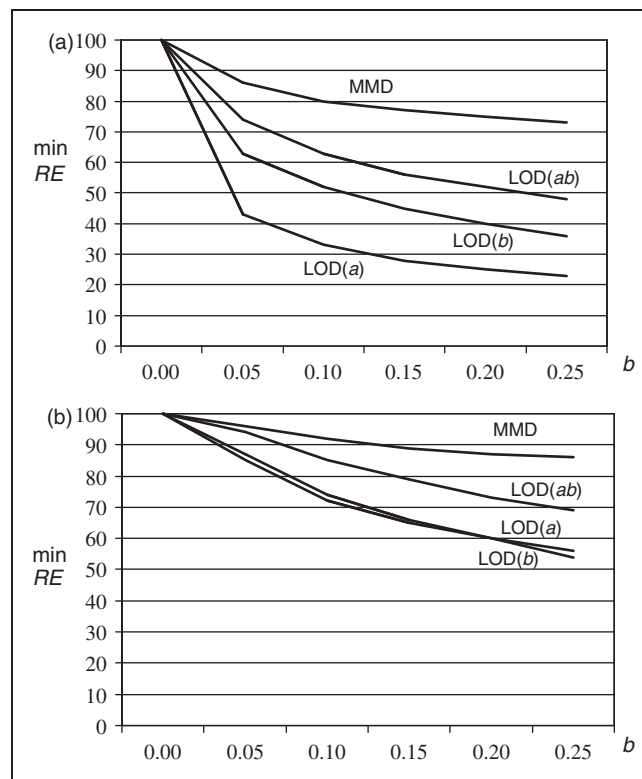
$$MMV = \frac{sn}{c + sn} = \frac{\sqrt{\frac{c}{s}} + 2\sqrt{\left(\frac{1-b}{b}\right)}}{2\sqrt{\frac{c}{s}} + 2\sqrt{\left(\frac{1-b}{b}\right)}} \quad (11)$$

which is decreasing both in  $b$  and in the cost ratio, reaching a minimum of 0.50 if  $b = 1$ . Of course,  $a = 0$  gives an impossible optimal design by equation (5). However, Equations (10) and (11) give some insight into the behaviour of MMD and MMV if  $a \approx 0$ . Restricting the parameter space to an interval  $[a, b]$  with  $0 < a < b < 1$ , gives a different MMD and a higher MMV. This is because the minimum  $RE$  of any design is attained at either of the  $ICC$  boundaries (Appendix) and so the





**Figure 2.** Plot of the relative efficiencies  $RE(n,a)$  and  $RE(n,b)$  of a design with cluster size  $n$  versus the locally optimal designs  $LOD(a)$  and  $LOD(b)$ , respectively, as a function of  $n$ , for  $a = 0.01$  and  $b = 0.10$  as lower and upper boundaries of the unknown ICC, assuming the budget and costs of Figure 1. MMD is at the intersection of both relative efficiencies.



**Figure 3.** Minimum relative efficiency (RE) in %, for the MMD and the  $LOD(b)$ ,  $LOD(a)$  and  $LOD(ab)$  design as a function of the ICC upper boundary  $b$ , given a lower boundary  $a = 0.001$  (panel a) or  $0.01$  (panel b), assuming the budget and costs of Figure 1. The minimum RE of the  $LOD(b)$  and  $LOD(ab)$  is obtained at ICC value  $a$ , the minimum RE of the  $LOD(a)$  at ICC value  $b$ , and the minimum RE of the MMD at both ICC values (Figure 2).

MMD is the design  $(n, K)$ , which maximizes the minimum of  $RE(n, a)$  and  $RE(n, b)$ , where both  $RE$ s are obtained from (7). Figure 2 shows how this leads to the MMD, and technical details are given in Appendix.

The results above assumed  $n, K > 0$ . In practice, there will be a lower bound larger than zero, and an upper bound, on  $n$  and  $K$ . Given the costs and budget, bounds on  $n$  imply bounds on  $K$  and *vice versa*. Hence, the derivation of the MMD can be repeated with a lower and upper bound for  $n$ , ignoring  $K$ . It can be shown that this gives a slightly different MMD which approaches (8) as the lower bound for  $n$  moves to zero and the upper bound for  $n$  becomes large. Further, its MMV is higher than according to (9).

## 5 Performance of the designs in terms of relative efficiency and efficiency

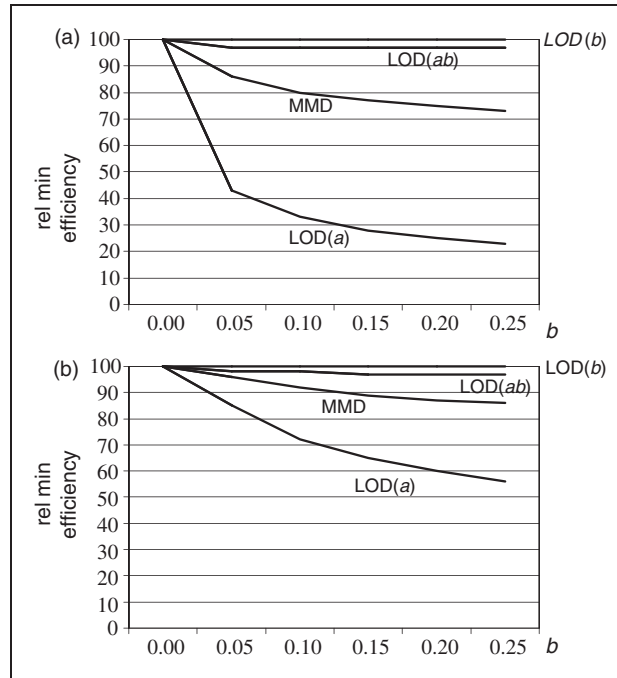
MMD maximizes the minimum  $RE$  and is thus robust against misspecification of the  $ICC$  if the MMV is high enough. If  $a = 0$ , it follows from (11) that the MMV decreases as  $b$  increases. With (9), it can be shown that this also holds if  $a > 0$ . To give an impression, Figure 3 plots the MMV against  $b$  for the cases  $a = 0.001$  and  $a = 0.01$ , and with  $b$  up to 0.25. To compare, the minimum  $RE$  is also plotted for the three locally optimal designs  $LOD(a)$ ,  $LOD(b)$  and  $LOD(ab)$ , where  $ab$  refers to the midpoint of the  $ICC$  range, that is, to  $(a + b)/2$ . Remember that, since  $\text{Var}(\hat{\beta}_1)$  is an increasing function of the  $ICC$ , the  $LOD(b)$  maximizes minimum efficiency instead of minimum  $RE$ , by minimizing the maximum possible  $\text{Var}(\hat{\beta}_1)$ . Similarly,  $LOD(a)$  minimizes the minimum possible  $\text{Var}(\hat{\beta}_1)$ , and  $LOD(ab)$  minimizes the expected  $\text{Var}(\hat{\beta}_1)$  if the  $ICC$  is symmetrically distributed over  $[a, b]$ . Figure 3 shows that MMD clearly outperforms LODs, and more so as  $b$  increases. This is logical, as the  $LOD(b)$  is very inefficient if  $\rho \approx a$  and the  $LOD(a)$  is very inefficient if  $\rho \approx b$ . For the same reason, the  $LOD(ab)$  design performs better than both  $LOD(a)$  and  $LOD(b)$ , although not as well as the MMD which by definition gives the highest minimum  $RE$ . This minimum  $RE$  or MMV depends on the  $ICC$  range and cost ratio  $c/s$ . For example, in Figure 3 where  $c/s = 20$ , the MMV is 0.80 if  $a = 0.001$  and  $b = 0.10$ , but 0.96 if  $a = 0.01$  and  $b = 0.05$ .

In terms of minimum efficiency (maximum variance) instead of minimum relative efficiency, things are different since the  $LOD(b)$  has the highest minimum efficiency. For each design, Figure 4 plots the relative minimum efficiency, that is, the ratio of its minimum efficiency to the minimum efficiency of the  $LOD(b)$ . Now,  $LOD(b)$  outperforms the MMD, especially if  $a = 0.001$  and  $b$  is large. Figure 4 also shows that the  $LOD(ab)$  design is almost as efficient as the  $LOD(b)$ , and that the  $LOD(a)$  is very inefficient. Note that the MMD and  $LOD(a)$  curves in Figure 4 are the same as in Figure 3, because these two designs have their minimum  $RE$  at the same  $ICC$  value where their efficiency is minimum, that is, at  $ICC = b$ .

The efficiency criteria in Figures 3 and 4 can be combined by taking their average, which is an example of a compound optimal design criterion.<sup>19</sup> The performance of all designs by this criterion is plotted in Figure 5, showing the MMD and  $LOD(ab)$  to perform equally well, and slightly better than the  $LOD(b)$  and much better than the  $LOD(a)$ . Finally, although all figures assumed a cost ratio  $c/s = 20$ , similar plots and the same conclusions were obtained with other cost ratios, as will be seen in Section 8.

## 6 Multicentre trials

Suppose that individual randomization is possible, but the study includes a large number of centres (e.g. general practices) to recruit sufficient persons and test for treatment  $\times$  centre interaction. We then have a multicentre trial, for which a simple model is



**Figure 4.** Relative minimum efficiency ( $\max \text{Var}(\hat{\beta}_1)$ ) in %, for the MMD and the LOD(b), LOD(a) and LOD(ab) design as a function of the ICC upper boundary  $b$ , given a lower boundary  $a = 0.001$  or  $0.01$ , assuming the budget and costs of Figure 1. The minimum efficiency is obtained at ICC value  $b$  for all designs. Relative minimum efficiency is relative to the LOD(b).

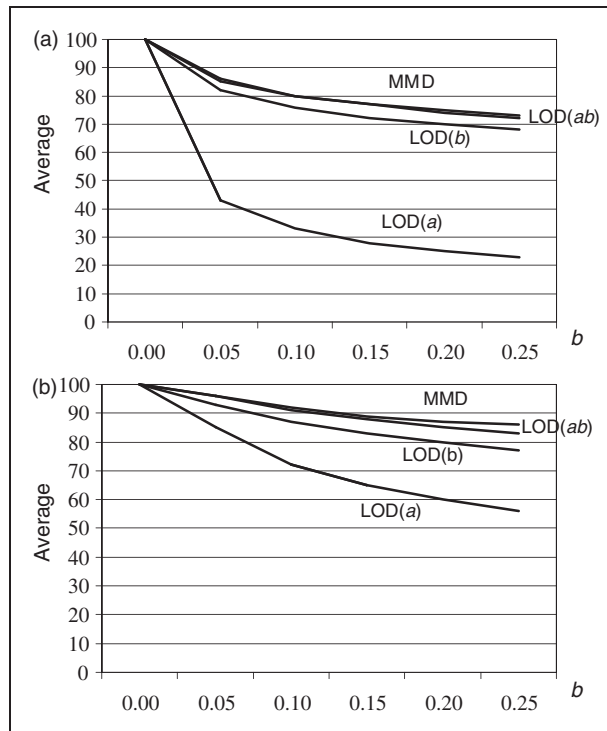
$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_{0j} + u_{1j} X_{ij} + e_{ij} \quad (12)$$

where  $Y_{ij}$  is the outcome for person  $i$  in cluster (centre)  $j$  and  $X_{ij}$  the treatment which is now coded as  $+0.5$  for treated and  $-0.5$  for control persons. The difference with model (1) for cluster randomized trials is the additional random effect  $u_{1j}$  with variance  $\sigma_1^2$ . This represents treatment  $\times$  centre interaction.

Assuming absence of covariance between  $u_{0j}$  and  $u_{1j}$  (i.e.  $\sigma_{01} = 0$ , implying homogeneity of variance between treatment arms), and 50:50 allocation within each centre, the variance of  $\hat{\beta}_1$  satisfies Equation (3) if the following definitions are used in (3):<sup>14,40</sup>

$$\sigma_Y^2 = \frac{1}{4}\sigma_1^2 + \sigma_e^2, \quad \rho = \frac{\frac{1}{4}\sigma_1^2}{\frac{1}{4}\sigma_1^2 + \sigma_e^2} \quad (13)$$

Hence, by replacing  $\sigma_0^2$  with  $\frac{1}{4}\sigma_1^2$ , Equation (3) is obtained, from which the LOD and MMD for cluster randomized trials were derived. Hence, Equations (5) and (8) for the LOD and MMD also hold for multicentre trials upon proper substitution. Using 0/1 coding for treatment in (3) gives homogeneity of variance if  $2\sigma_{01} = -\sigma_1^2$  and a reparametrization of (13), but leads to the same LOD and MMD. For the case of heterogeneous variance, see Appendix A2 in the study of Candel *et al.*<sup>40</sup>



**Figure 5.** Average of minimum relative efficiency and relative minimum efficiency in %, for the MMD and the LOD( $b$ ), LOD( $a$ ) and LOD( $ab$ ) design as a function of the ICC upper boundary  $b$ , given a lower boundary  $a = 0.001$  or  $0.01$ , assuming the budget and costs of Figure 1.

## 7 Binary outcomes

Binary outcomes in a cluster randomized trial can be analysed by mixed logistic regression, among others. The model of interest then is:

$$LN\left(\frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)}\right) = \beta_0 + \beta_1 X_j + u_{0j} \quad (14)$$

where  $u_{0j}$  is a random cluster effect with mean zero and variance  $\sigma_0^2$ . An expression for the variance of the treatment effect estimator has been derived,<sup>8</sup> which can be rewritten into:

$$\text{Var}(\hat{\beta}_1) = \frac{4}{K} \left[ \sigma_0^2 + \frac{\delta^2}{n} \right]$$

where

$$\delta^2 = \frac{1}{2} \left( \frac{1}{\pi_c(1 - \pi_c)} + \frac{1}{\pi_t(1 - \pi_t)} \right) \quad (15)$$

Here,  $\delta^2$  is the person-level variance on the log odds scale, similarly to  $\sigma_e^2$  in the mixed linear model (1), and  $\pi_c$  and  $\pi_t$  the probabilities  $P(Y_{ij} = 1)$  for the control and treated arms, respectively,

if  $u_{0j} = 0$ . This equation is based on first-order Marginal Quasi-Likelihood (MQL1) estimation, which underestimates  $\text{Var}(\hat{\beta}_1)$ . A fairly unbiased method is second-order Penalized Quasi-Likelihood (PQL2), but no closed forms like Equation (15) can be derived for PQL2. However, simulation studies show that multiplying  $\text{Var}(\hat{\beta}_1)$  as computed with (15) with a factor 1.25 is a safe correction for sample size planning<sup>8,26</sup> when using PQL2.

What is important here is that (15) can be rewritten into (3) if we define  $\delta^2 = \sigma_e^2$ . Since all equations for the LOD and MMD were derived from (3), it follows that the results in this article also hold for binary outcomes if we replace  $\sigma_e^2$  with  $\delta^2$  as defined in (15). Finally, the results can be extended to multicentre trials with binary outcome by performing two substitutions in (3) simultaneously: replacing  $\sigma_0^2$  with  $\sigma_1^2/4$  and replacing  $\sigma_e^2$  with  $\delta^2$ . In both cases, cluster randomized and multicentre trial,  $\text{Var}(\hat{\beta}_1)$  and by Equation (3) also the number of clusters  $K$  and budget  $B$ , must finally be multiplied with the correction factor 1.25 for using PQL2 instead of MQL1.

## 8 Application and comparison with published trials

The theory of Sections 4 and 5 will now be applied to compute optimal and efficient designs for two realistic  $ICC$  ranges,  $[0.001, 0.05]$  and  $[0.01, 0.10]$ , and using as cost ratios  $c/s = 5, 20$  and  $50$ , based on a reported cost ratio of  $c/s = 26$  in a cluster randomized smoking prevention study in primary schools,<sup>1,41</sup> as no information on sampling costs in other trials was found. For each combination of  $ICC$  range and cost ratio, the cluster sizes of the MMD, LOD( $a$ ), LOD( $b$ ) and LOD( $ab$ ) were computed as well as their minimum relative efficiency and relative minimum efficiency. The number of clusters per design then follows from cluster size and budget (Equation (4)). Designs and efficiencies are summarized in Table 2.

Table 2 shows several things. First, for every cost ratio and  $ICC$  range, the designs can be ordered as LOD( $a$ ), MMD, LOD( $ab$ ), LOD( $b$ ) in terms of cluster size. Second, the MMD and LOD( $ab$ ) are both

**Table 2.** MMD and LOD at  $ICC$  values  $a, b$  and  $(a + b)/2$  for cost ratios  $c/s = 5, 20, 50$  and  $ICC$  ranges  $[a, b] = [0.001, 0.05]$  and  $[0.01, 0.10]$ .

Cost ratio	Design	ICC range 0.001–0.05			ICC range 0.01–0.10		
		Cluster size, $n$	Min RE	Relative minimum efficiency	Cluster size, $n$	Min RE	Relative minimum efficiency
5	LOD( $a$ )	70.7	0.45	0.45	22.3	0.72	0.72
	MMD	19.7	0.90	0.90	11.4	0.93	0.93
	LOD( $ab$ )	13.8	0.83	0.97	9.3	0.89	0.97
	LOD( $b$ )	9.8	0.75	1.00	6.7	0.80	1.00
20	LOD( $a$ )	141.4	0.43	0.43	44.5	0.72	0.72
	MMD	43.5	0.86	0.86	24	0.92	0.92
	LOD( $ab$ )	27.7	0.74	0.97	18.5	0.85	0.98
	LOD( $b$ )	19.5	0.63	1.00	13.4	0.74	1.00
50	LOD( $a$ )	223.5	0.44	0.44	70.4	0.75	0.75
	MMD	74.8	0.83	0.83	39.5	0.92	0.92
	LOD( $ab$ )	43.7	0.67	0.97	29.3	0.83	0.98
	LOD( $b$ )	30.8	0.55	1.00	21.2	0.72	1.00

Notes: LOD( $ab$ ) is the LOD at the midpoint  $(a + b)/2$  of the  $ICC$  range. Minimum relative efficiency (min RE) is attained at  $\rho = a$  or at  $\rho = b$  for every design. Relative minimum efficiency is the RE at  $\rho = b$  with LOD( $b$ ) as a reference.

quite efficient, with the MMD beating the  $\text{LOD}(ab)$  on minimum  $RE$  and the  $\text{LOD}(ab)$  beating the MMD on relative minimum efficiency. The  $\text{LOD}(b)$  is of course even better on relative minimum efficiency, but worse on minimum  $RE$ . The  $\text{LOD}(a)$  is worst on both criteria. Third, an increasing cost ratio leads to larger clusters and an increasing  $ICC$  leads to smaller clusters, for all designs. Finally, the optimal cluster size is between 10 and 50 for all cost ratios and  $ICC$  ranges considered, except one MMD case and most  $\text{LOD}(a)$  cases.

What do these results imply for the actual study designs in Table 1? Cluster sizes much smaller than 10 and much larger than 50 do occur in various studies, but are far from optimal for the cost ratios and  $ICC$  ranges in Table 2. Cluster sizes smaller than 10 are efficient only at cost ratios below 5 and/or  $ICC$  ranges with an upper bound above 0.10. For instance, Equation (5) gives  $n = 4$  as the  $\text{LOD}(ab)$  at a cost ratio  $c/s = 1$  if the  $ICC$  range is  $[0.01, 0.10]$ , or at  $c/s = 2$  if the  $ICC$  range is  $[0.01, 0.20]$ . Likewise, cluster sizes larger than 50 are efficient only at cost ratios above 50 and/or  $ICC$  ranges with  $b < 0.05$  or  $a = 0$ . For example,  $n = 100$  is the  $\text{LOD}(ab)$  if the cost ratio is 100 and the  $ICC$  range is  $[0, 0.02]$ .

This is not to say that the designs in Table 1 were ill-chosen, since a design choice also depends on other considerations than costs, like feasibility. Sometimes, it is difficult to include more general practices. A sample size of  $n > 50$  per practice can then offer some compensation, although Equation (3) shows that this compensation is ignorable if  $n\rho$  is much larger than 1. Sometimes, it is difficult to include more patients per practice within the time frame of the study so that  $n < 10$ . This can then be compensated by including more practices. However, the present results do indicate which samples sizes are usually optimal or at least efficient. A researcher may first choose an optimal or efficient design based on the present tables and equations. If that design is not feasible, one may then choose the nearest (in terms of cluster size and number of clusters) design which is feasible.

## 9 Discussion

Starting from a simple model for the treatment effect on a quantitative outcome in a cluster randomized trial, a practical expression was given for the optimal design (number of clusters, number of persons per cluster) under budget and cost constraints. It was seen that the optimal design depends not only on the sampling budget and costs, but also on the  $ICC$  which is rarely known in study planning. This problem was handled using the Maximin procedure, leading to a design which is robust against misspecification of the  $ICC$  in either of two senses: Having the highest minimum relative efficiency ( $RE$ ) compared with the optimal design (MMD), or having the highest minimum efficiency ( $\text{LOD}(b)$ ), where the minimum is taken over the  $ICC$  range. It was shown that the MMD has a high minimum  $RE$  for realistic  $ICC$  ranges and cost ratios and can thus be recommended for practical use in many cases. It was also shown that the  $\text{LOD}(ab)$ , which is the  $\text{LOD}$  at the midpoint of the  $ICC$  range, is a good alternative to the  $\text{LOD}(b)$  as it has almost the same minimum efficiency and a higher minimum  $RE$ . Comparing MMD with  $\text{LOD}(ab)$ , the MMD is better on minimum  $RE$ , the  $\text{LOD}(ab)$  is better on minimum efficiency, and they perform equally well on the average of both criteria. Finally, the  $\text{LOD}(a)$  has a low minimum efficiency (high maximum  $\text{Var}(\hat{\beta}_1)$ ) and a low minimum  $RE$ , so the  $\text{LOD}(a)$  is never to be recommended. It was also shown how the theory applies to multicentre trials with treatment  $\times$  centre interaction and to binary outcomes, and the theory was illustrated with data from published cluster randomized trials.

There are several limitations and possible extensions to this study, which will now be touched upon. First of all, we assumed equal cluster sizes, which are not very realistic. However, as discussed in Section 1, current literature provides approximations for the relative efficiency of varying *versus* equal cluster sizes based on  $CV$  of cluster sizes, which can be used to correct the total number of clusters  $K$  as obtained with the equations in this article (for details, see Van Breukelen *et al.*<sup>14</sup> and Candel and

Van Breukelen<sup>26</sup>). As a second limitation, the mixed model of Equation (1) does not include any covariates. Ignoring here possible treatment  $\times$  covariate interaction, including a covariate has two potential effects. On the one hand, it reduces unexplained outcome variance and thereby increases power and precision of treatment effect estimation. This holds for standard clinical trials and also for cluster randomized trials, as a cluster-level covariate reduces cluster-level variance  $\sigma_0^2$  and a person-level covariate reduces person-level variance  $\sigma_e^2$ .<sup>9</sup> On the other hand, treatment–covariate correlation increases  $\text{Var}(\hat{\beta}_1)$ . However, cluster randomization ensures that treatment–covariate correlation goes to zero for a cluster-level covariate as the number of clusters  $K$  increases, and that the correlation is zero for a person-level covariate. Likewise, individual randomization as in multicentre trials ensures that treatment–covariate correlation goes to zero for a person-level covariate as the total number of persons  $nK$  increases and is zero for a cluster-level covariate.<sup>9</sup> Ignoring covariates in sample size planning is therefore a safe approach, while taking them into account may save money by reducing outcome variance at each level of the design. A third limitation of the present work is that, compared with Bayesian optimal design, MMD has the drawback of considering all *ICC* values in the assumed range equally important, irrespective their prior probability. On the other hand, Maximin also has several advantages over Bayesian design, notably ease of use and invariance of the design to monotonic transformation of the optimality criterion. MMDs are easily computed with the equations in this article and do not require the construction of a prior distribution for the *ICC*. Further, using efficiency or variance as the design criterion gives the same MMD, but may give different Bayesian designs since efficiency and variance are non-linearly related to each other. In fact, Bayesian optimal design literature frequently uses as criterion the log efficiency,<sup>36,38</sup> which may give yet another Bayesian optimal design. A systematic comparison between Bayesian and MMD under various priors and cost constraints is therefore desirable. A final limitation is that we assumed homogeneity of variance and sampling costs between treatment arms, which may not always be realistic.<sup>42</sup> For the classical RCT without nesting, it has been shown<sup>43</sup> that the optimal treatment-to-control allocation ratio increases as the treatment-to-control variance ratio increases or as the treatment-to-control cost ratio decreases. Ongoing work by the present authors gave a similar result for the LOD of a cluster randomized trial with heterogeneous variances and costs. However, as the two cluster-level variances and two person-level variances are unknown, MMD over a three-dimensional space of variance ratios is needed, which is beyond the present scope.

To conclude, cost effectiveness, a widely accepted criterion for evaluating medical treatments and public health programs, can also be used to evaluate clinical trials. Cost-effective design of cluster randomized and multicentre trials increases the number of trials that can be run for the same budget. However, this requires knowledge of the *ICC* range and of the cost ratio. It is therefore important that publications of cluster randomized and multicentre trials provide information not only on *ICCs*, but also on the costs per included cluster/centre and per person, to improve the design of future trials.

## References

1. Ausems M, Mesters I, Van Breukelen G and De Vries H. Short-term effects of a randomised computer-based out-of-school smoking prevention trial aimed at Dutch elementary schoolchildren. *Prev Med* 2002; **34**(6): 581–589. DOI 10.1006/pmed.2002.1021.
2. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000; **9**: 81–94.
3. Moerbeek M, Van Breukelen GJP and Berger MPF. A comparison between traditional methods and multilevel regression for the analysis of multicentre intervention studies. *J Clin Epidemiol* 2003; **56**(4): 341–350. DOI 10.1016/S0895-4356(03)00007-6.
4. Senn S. Some controversies in planning and analyzing multicentre trials. *Stat Med* 1998; **17**(15–16): 1753–1765. DOI 10.1002/(sici)1097-0258(19980815/30).
5. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychol Methods* 1997; **2**(2): 173–185.
6. Raudenbush SW and Liu X. Statistical power and optimal design for multisite trials. *Psychol Methods* 2000; **5**(2): 199–213. DOI 10.1037/1082-989X.5.2.199.



7. Moerbeek M, Van Breukelen GJP and Berger MPF. Design issues for experiments in multilevel populations. *J Educ Behav Stat* 2000; **25**(3): 271–284.
8. Moerbeek M, Van Breukelen GJP and Berger MPF. Optimal experimental design for multilevel logistic models. *The Statistician* 2001; **50**(1): 17–30.
9. Moerbeek M, Van Breukelen GJP and Berger MPF. Optimal experimental designs for multilevel models with covariates. *Commun Stat Theory Methods* 2001; **30**(12): 2683–2697. DOI 10.1081/STA-100108453.
10. Liu X. Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *J Educ Behav Stat* 2003; **28**(3): 231–248.
11. Headrick TC and Zumbo BD. On optimizing multi-level designs: power under budget constraints. *Austr N Z J Stat* 2005; **47**(2): 219–229. DOI 10.1111/j.1467-842x.2005.00386.x
12. Lancaster GA, Campbell MJ, Eldridge S, Farrin A, Marchant M, Muller S, et al. Trials in primary care: statistical issues in the design, conduct and evaluation of complex interventions. *Stat Methods Med Res* 2010; **19**: 349–377. DOI 10.1177/0962280209359883.
13. Ankenman BE, Aviles AI and Pinheiro JC. Optimal designs for mixed-effects models with two random nested factors. *Stat Sin* 2003; **13**(2): 385–401.
14. Van Breukelen GJP, Candel MJJM and Berger MPF. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Stat Med* 2007; **26**(13): 2589–2603. DOI 10.1002/sim.2740.
15. Campbell MJ, Donner A and Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Stat Med* 2007; **26**(1): 2–19. DOI 10.1002/sim.2731.
16. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S and Campbell MJ. Patterns of intracluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004; **57**: 785–794. DOI 10.1016/j.jclinepi.2003.12.013.
17. Eldridge SM, Ashby D, Feder GS, Rudnicka AR and Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials* 2004; **1**: 80–90. DOI 10.1191/1740774504cn006rr.
18. Atkinson AC, Donev AN and Tobias RD. *Optimum experimental designs, with SAS*. Oxford, UK: Oxford University Press, 2007.
19. Berger MPF and Wong WK. *An introduction to optimal designs for social and biomedical research*. Chichester, UK: Wiley, 2009.
20. Manatunga AK, Hudgens MG and Chen S. Sample size estimation in cluster randomized studies with varying cluster size. *Biomet J* 2001; **43**(1): 75–86. DOI 10.1002/1521-4036(200102).
21. Jung S, Kang S and Ahn CW. Sample size calculations for clustered binary data. *Stat Med* 2001; **20**(13): 1971–1982. DOI 10.1002/sim.846.
22. Lake S, Kammann E, Klar N and Betensky R. Sample size re-estimation in cluster randomized trials. *Stat Med* 2002; **21**(10): 1337–1350. DOI 10.1002/sim.1121.
23. Normand ST and Zou KH. Sample size considerations in observational health care quality studies. *Stat Med* 2002; **21**(3): 331–345. DOI 10.1002/sim.1020.
24. Eldridge SM, Ashby D and Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol* 2006; **35**(5): 1292–1300. DOI 10.1093/ije/dyl129.
25. Taljaard M, Donner A and Klar N. Accounting for expected attrition in the planning of community intervention trials. *Stat Med* 2007; **26**(13): 2615–2628. DOI 10.1002/sim.2733.
26. Candel MJJM and Van Breukelen GJP. Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. *Stat Med* 2010; **29**(14): 1488–1501. DOI 10.1002/sim.3857.
27. Ouwens JNM, Tan FES and Berger MPF. Maximin D-optimal design for longitudinal mixed effects models. *Biometrics* 2002; **58**(4): 735–741.
28. Winkens B, Schouten HJA, Van Breukelen GJP and Berger MPF. Optimal designs for clinical trials with second-order polynomial treatment effects. *Stat Methods Med Res* 2007; **16**(6): 523–537. DOI 10.1177/0962280206071847.
29. Maus B, Van Breukelen GJP, Goebel R and Berger MPF. Robustness of optimal design of fMRI experiments with application of a genetic algorithm. *NeuroImage* 2010; **49**: 2433–2443. DOI 10.1016/j.neuroimage.2009.10.004.
30. Campbell MK, Mollison J and Grimshaw J. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001; **20**: 391–399.
31. Smeeth L and Siu-Woon E. Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC trial of the assessment and management of older people in the community. *Control Clin Trials* 2002; **23**(4): 409–421. DOI 10.1016/S0197-2456(02)00208-8.
32. Elley CR, Kerse N, Chondros P and Robinson E. Intraclass correlation coefficients from three cluster randomized controlled trials in primary and residential health care. *Austr N Z J Public Health* 2005; **29**: 461–467.
33. Parker DR, Evangelou E and Eaton CB. Intraclass correlation coefficients for cluster randomized trials in primary care: the cholesterol education and research trial CEART. *Contemp Clin Trials* 2005; **26**: 260–267. DOI 10.1016/j.cct.2005.01.002.
34. Littenberg B and MacLean CD. Intra-cluster correlation coefficients in adults with diabetes in primary care practices: the Vermont Diabetes Information System field survey. *BMC Med Res Methodol* 2006; **6**: 20. DOI 10.1186/1471-2288-6-20.
35. Donner A, Birkett N and Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol* 1981; **114**: 906–914.
36. Chaloner K and Verdinelli L. Bayesian experimental design: a review. *Stat Sci* 1995; **10**(3): 273–304.
37. Turner RM, Thompson SG and Spiegelhalter DJ. Prior distributions for the intracluster correlation coefficient, based on multiple previous estimates, and their application in cluster randomized trials. *Clin Trials* 2005; **2**: 108–118. DOI 10.1191/1740774505cn072oa.
38. Khuri AI, Mukherjee B, Sinha BK and Ghosh M. Design issues for generalized linear models: a review. *Stat Sci* 2006; **21**(3): 376–399. DOI 10.1214/088342306000000105.
39. Rotondi MA and Donner A. Sample size estimation in cluster randomized educational trials: An empirical Bayes approach. *J Educ Behav Stat* 2009; **34**(2): 229–237. DOI 10.3102/1076998609332756.
40. Candel MJJM, Van Breukelen GJP, Kotova L and Berger MPF. Optimality of unequal cluster sizes in multilevel studies with realistic sample sizes. *Commun Stat Simul Comput* 2008; **37**(1): 222–239. DOI 10.1080/03610910701724052.
41. Moerbeek M, Van Breukelen GJP, Berger MPF and Aulsems M. Optimal sample sizes in experimental designs with individuals nested within clusters. *Understand Stat* 2003; **2**(3): 151–175.

42. Roberts C and Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clin Trials* 2005; **2**: 152–162.
43. Schouten HJA. Sample size formula with a continuous outcome for unequal group sizes and unequal variances. *Stat Med* 1999; **18**: 87–91.

## I APPENDIX

### I.1 Derivation of the MMD and its minimum RE

The proof consists of the following steps: First, show that the relative efficiency  $RE(n, \rho)$  of design  $(n, K)$  against  $LOD(\rho)$  is a single-peaked function of  $\rho$  with a maximum of 1 at the  $\rho$ -value for which the design  $(n, K)$  is itself optimal by Equation (5), and a minimum at either of the boundaries  $a$  and  $b$  for  $\rho$ . Second, show that  $RE(n, a)$  and  $RE(n, b)$  are single-peaked functions of cluster size  $n$  and that the minimum  $RE$  is maximized by the design satisfying  $RE(n, a) = RE(n, b)$ . Third, find this design.

#### Step 1: Proof that the $RE$ of design $(n, K)$ is minimal at $\rho = a$ or at $\rho = b$

From Equations (3) to (6), it follows that the  $RE$  of a design  $(n, K)$  versus the  $LOD(\rho)$  is:

$$RE(n, \rho) = \frac{g(\rho)}{[(n-1)\rho + 1]} \times \frac{n}{c + sn} \quad (A1)$$

where  $g(\rho) = \left( \sqrt{\rho c} + \sqrt{(1-\rho)s} \right)^2$  as in Equation (6), and  $K$  and  $B$  cancel out due to Equation (4), making the  $RE$  a function of cluster size  $n$  and ICC  $\rho$  only, given the sampling costs  $c$  per cluster and  $s$  per person upon which  $g(\rho)$  depends.

Taking the partial derivative with respect to  $\rho$  gives:

$$\begin{aligned} \frac{\partial RE(n, \rho)}{\partial \rho} = & \frac{\sqrt{g(\rho)}}{[(n-1)\rho + 1]^2} \times \frac{n}{c + sn} \times \\ & \left\{ \left[ \left( \frac{c}{\rho} - \sqrt{\frac{s}{1-\rho}} \right) ((n-1)\rho + 1) \right] - \left[ \left( \sqrt{\rho c} + \sqrt{(1-\rho)s} \right) (n-1) \right] \right\} \end{aligned}$$

of which the first two factors are always positive. Setting the third factor to zero gives:

$$\rho = \frac{c}{c + sn^2}, \text{ or equivalently, } n = \sqrt{\left( \frac{1-\rho}{\rho} \right) \left( \frac{c}{s} \right)} \quad (A2)$$

which is the  $LOD$  of Equation (5). Denoting the  $\rho$  satisfying (A2) as  $\rho_0$  it can be verified that  $\frac{\partial RE(n, \rho)}{\partial \rho} > 0$  if  $a \leq \rho < \rho_0$ , and  $\frac{\partial RE(n, \rho)}{\partial \rho} < 0$  if  $\rho_0 < \rho \leq b$ , and so the  $RE$  attains its maximum of 1 at  $\rho_0$ , and its minimum either at  $\rho = a$  or at  $\rho = b$ , assuming  $\rho_0 \in [a, b]$ . Designs which are locally optimal for some  $\rho_0 \leq a$  have their maximum  $RE$  at  $\rho = a$  and their minimum  $RE$  at  $\rho = b$ . Designs which are locally optimal for some  $\rho_0 \geq b$  have their maximum  $RE$  at  $\rho = b$  and their minimum  $RE$  at  $\rho = a$ .

#### Step 2: Proof that min $RE$ is maximized by the design satisfying $RE(n, a) = RE(n, b)$

Inserting  $\rho = a$  in Equation (A1) and taking the partial derivative with respect to  $n$  gives:

$$\begin{aligned} \frac{\partial RE(n, a)}{\partial n} = & \frac{g(a)}{\{[(n-1)a + 1][c + sn]\}^2} \\ & \times \{[(n-1)a + 1][c + sn] - [na(c + sn) + n((n-1)a + 1)s]\} \end{aligned}$$

The first factor is always positive. Rewriting the second factor shows the derivative to be zero at  $n_a = \sqrt{\frac{(1-a)c}{as}}$ , which is the LOD( $a$ ) as Equation (5) shows, giving  $RE(n,a) = 1$  if  $n = n_a$ . It can be verified that  $\frac{\partial RE(n,a)}{\partial n} > 0$  if  $n < n_a$  and  $\frac{\partial RE(n,a)}{\partial n} < 0$  if  $n > n_a$ , and so  $RE(n,a)$  is a single-peaked function of cluster size  $n$  with maximum of 1 at  $n = n_a$ . Similarly, we obtain  $n_b = \sqrt{\frac{(1-b)c}{bs}}$  as the solution of  $\frac{\partial RE(n,b)}{\partial n} = 0$ , giving  $RE(n,b) = 1$  as maximum at  $n = n_b$ . Further, since  $b > a$ , it follows that  $n_b < n_a$ . Hence, plotting both  $RE(n,a)$  and  $RE(n,b)$  as a function of cluster size  $n$ , we have that both increase if  $n < n_b$ , that  $RE(n,b)$  already decreases whereas  $RE(n,a)$  still increases if  $n_b < n < n_a$ , and that both decrease if  $n > n_a$ . Since  $RE(n,b)$  and  $RE(n,a)$  both have a maximum value of 1, attained at  $n = n_b$  and at  $n = n_a$ , respectively, it follows that they intersect at some value  $n = n_0$  which satisfies  $n_b < n_0 < n_a$  (Figure 2). Denote the common  $RE$  value at the intersection by  $RE_0$ . Since  $RE(n,a)$  is still increasing in  $n$  whereas  $RE(n,b)$  is already decreasing at  $n = n_0$ , it then follows that  $RE(n,a) < RE_0$  for all  $n < n_0$  and that  $RE(n,b) < RE_0$  for all  $n > n_0$ . In short,  $\min\{RE(n,a), RE(n,b)\} < RE_0$  for all  $n \neq n_0$ , and so  $n = n_0$  is the MMD and its minimum  $RE$  value (MMV) is  $RE_0$ .

### Step 3: Derivation of the MMD

To find both the MMD  $n_0$  and its MMV  $RE_0$ , observe that:

$$RE(n,a) = RE(n,b) \Rightarrow g(a)[(n-1)b+1] = g(b)[(n-1)a+1] \quad (\text{A3})$$

which gives as the only possible solution:

$$n_0 = \frac{(b-1)g(a) - (a-1)g(b)}{bg(a) - ag(b)} \quad (\text{A4})$$

To show that  $n_0 > 0$ , it is sufficient to show that denominator and numerator of (A4) are both positive if  $b > a$ : Replacing, in the denominator of Equation (A4),  $g(a)$  and  $g(b)$  with their definition in Equation (6) gives after some rewriting:

$$bg(a) - ag(b) = s(b-a) + 2\sqrt{cs} \times [b\sqrt{a(1-a)} - a\sqrt{b(1-b)}]$$

The part between the [ ] sign can be verified to be positive for all  $b > a$ , and all other terms are likewise positive. By similar rewriting the numerator of (A4) can be shown to be

$$(b-1)g(a) - (a-1)g(b) = c(b-a) + 2\sqrt{cs} \times [(1-a)\sqrt{b(1-b)} - (1-b)\sqrt{a(1-a)}]$$

The part between the [ ] is again positive for all  $b > a$ , as are all other terms.

The MMV of the MMD is finally obtained by inserting (A4) into (7), giving Equation (9).